A LAYERED STEREO ALGORITHM USING IMAGE SEGMENTATION AND GLOBAL VISIBILITY CONSTRAINTS

Michael Bleyer and Margrit Gelautz

Interactive Media Systems Group, Vienna University of Technology
Favoritenstrasse 9-11/188/2, A-1040 Vienna, Austria, [bleyer, gelautz]@ims.tuwien.ac.at

ABSTRACT

We propose a new stereo algorithm which uses colour segmentation to allow the handling of large untextured regions and precise localization of depth boundaries. Each segment is modelled as a plane. Robustness of the depth representation is achieved by the use of a layered model. Layers are extracted by mean-shift-based clustering of depth planes. For layer assignment a global cost function is defined. The quality of the disparity map is measured by warping the reference image to the second view and comparing it with the real image. Z-buffering enforces visibility and allows the explicit detection of occlusions. An efficient greedy algorithm searches for a local minimum of the cost function. Layer extraction and assignment are alternately applied. Results obtained for benchmark and self-recorded images indicate that the proposed algorithm can compete with the state-of-the-art.

1. INTRODUCTION

In our work we propose a stereo algorithm that represents the scene as a collection of planar layers. As a result we obtain piecewise smooth surface reconstructions and real-valued disparity estimates providing a high precision. Our algorithm explicitly addresses major problems arising in stereo computation. Large untextured regions are handled by applying colour segmentation to the reference image. Smoothness inside the derived segments is enforced by the use of a planar model representing each segment's disparity. Colour segmentation also allows the accurate localization of depth discontinuities. Occlusions in the reference and in the second view are detected and handled in a layer assignment step. Furthermore, we model smoothness across segments.

For a review of prior work we refer to [1], who give an extensive survey on recent stereo algorithms. In the following, we summarize the works most relevant to our approach. A model of planar layers for stereo was used in [2]. A surface fitting and a surface assignment step are alternately applied until convergence. For assigning pixels to surfaces a graph-based method is used. The work was extended in [3] with the most significant difference being the strictly symmetrical treatment of input images. In [4] the mean-shift algorithm was used for the extraction of planar layers in motion. Among prior work, the most similar to ours is the approach by Tao et al. [5]. We share the ideas of image warping for measuring the quality of a depth solution and hypothesizing depth from neighbouring segments. In contrast to Tao et al., we use a layered representation providing more robust depth solutions. A different cost function accounts for occlusions in both views and smoothness across segments. Furthermore, we compute new planar models throughout the whole process and achieve a higher amount of efficiency in our layer assignment step.

This work was supported by the Austrian Science Fund (FWF) under project P15663.

2. ALGORITHM

In the following, we describe the algorithm's building blocks and then show their integration into the overall algorithm. The input to our algorithm is formed by two epipolar rectified images. Throughout this paper, we refer to regions of homogeneous colour as $segments$. $Layers$ are groups of segments that can be approximated by the same planar equation.

2.1. Colour segmentation and planar model

We assume that for regions of homogeneous colour the disparity varies smoothly and depth discontinuities coincide with the boundaries of those regions [4, 5, 6], which holds true for most natural scenes. This assumption is incorporated by applying colour segmentation to the reference image and by using a planar model to represent the disparity inside the derived segments. It is generally safer to oversegment the image to ensure that this assumption is met. For segmentation we use the algorithm proposed in [7]. The resulting colour segmentation for a well-known stereo pair from the University of Tsukuba is shown in figure 1c.

We compute a sparse initial disparity map using a window-based method which exploits the results of the image segmentation. We calculate the sum of absolute differences

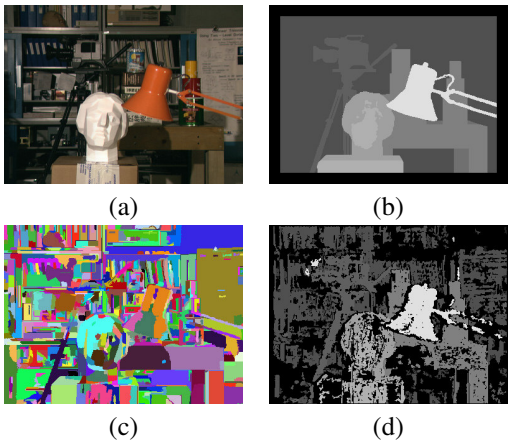


Fig. 1. Colour segmentation and initial disparity map. (a) Left image. (b) Ground truth provided with image pair. (c) Computed colour segmentation. (d) Computed initial disparity map. Invalid points are coloured black.

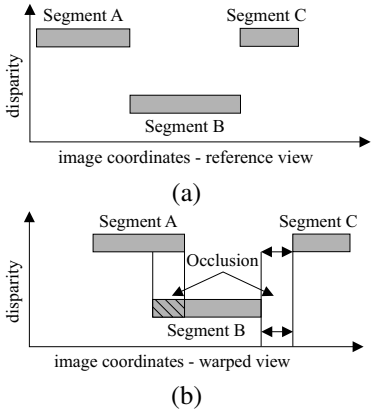


Fig. 2. The warping operation. (a) Segments and corresponding depth in the reference view. (b) Segments warped to the second view according to their depth planes.

(SAD) with a small (3×3) window. Cross validation is applied to filter out occluded points and areas of low texture, where disparity estimates tend to be unreliable. Similar to [6], we label those segments that have a density of valid points $> 50\%$ as reliable. For reliable segments we reduce the search scope based on the minimum and maximum disparity of valid points inside the segment. The reduction of the search scope helps to propagate good disparity inside the segment. The process can optionally be repeated with increasing window sizes, leaving the already found valid points unchanged. Using larger windows performs better in less-textured regions, but also intensifies the well-known foreground fattening effect. Figure 1d shows the initial disparity map calculated for the Tsukuba image using only a 3×3 window. A robust version of the method of least squared errors is then used to derive a plane equation for each segment. The plane is thereby fitted to all valid points of the initial disparity map inside the segment. The computed planes will be used in the layer extraction step.

2.2. Layer extraction

One single surface of the real world will usually be divided into several segments by applying colour segmentation. However, for segments of the same surface the planar models should be very similar, as long as the surface can be well approximated as a plane. Following this idea, we project each segment into a 5-dimensional feature space, consisting of 3 plane parameters and 2 spatial parameters representing the center of gravity. Segments of the same surface should then naturally build a cluster in this feature space. We use a modified version of the mean-shift algorithm [8] for extracting clusters. Members of the same cluster build a layer. For deriving a layer's plane equation we use the initial disparity map. Robust plane fitting is applied to the valid points of all segments belonging to the layer.

2.3. Layer assignment

In a hypothesis testing framework we assign each segment to a layer. The optimality of the assignment is measured by a global cost function. The basic idea behind the cost function is that if we warp the reference image to the second view according to the correct disparity map, the warped image should be very similar to the real image from this viewpoint. Translated to our cost function, we calculate the colour dissimilarity between the warped and the real view for all pixels visible in the warped image. The implemented warping operation is illustrated in figure 2. Visibility is naturally enforced using a Z-buffer that represents the second view. If a Z-buffer cell contains more than one pixel, only the pixel with the highest disparity is visible and the others are occluded in the second view. Empty Z-buffer cells represent occlusions in the reference image. In our cost function we have to penalize every detected occlusion, since otherwise declaring all pixels as occluded would be a trivial optimum. The last term accounts for modelling smoothness across segments. We introduce a discontinuity penalty that is given if two neighbouring pixels (in 4-connectivity) are assigned to different layers in the reference image. Summarising the above, we define the cost function

$$C = \sum_{p \in V} d(W(p), R(p)) + N_{occ}\lambda_{occ} + N_{disc}\lambda_{disc} \quad (1)$$

with V being the set of visible pixels, $d(W(p), R(p))$ being the dissimilarity function of the pixel p in the warped image $W(p)$ and in the real second view $R(p)$, which is implemented as the summed up absolute differences of RGB

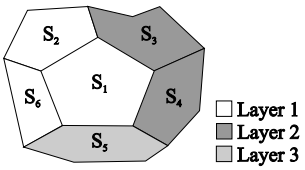


Fig. 3. Hypothesis testing. The segment S_1 has 5 neighbouring segments assigned to the layers 1, 2 and 3 which we call the neighbouring layers of S_1. We avoid testing layer 1 on S_1, since this is the current assignment. The layer hypotheses of layer 2 and 3 need to be checked. We point out that although there may be a large number of neighbouring segments, the layer neighbourhood is usually very small, which is a major argument for the algorithm's efficiency.

values, N_{occ} and N_{disc} being the number of detected occlusions and discontinuities and λ_{occ} and λ_{disc} are constant penalties for occlusion and discontinuity.

Unfortunately finding the layer assignment with minimum value for C is np-complete. We therefore use a greedy search strategy to find a local optimal solution. In the initial solution we use the layer assignment derived in the clustering step. We then hypothesize a segment's depth from a neighbouring layer. In hypothesis testing we replace the segment's current plane with the plane equation of the neighbouring layer and evaluate the cost function. The main idea is to propagate correct depth to untextured and occluded regions. For each segment we therefore test the plane equations of all neighbouring layers as shown in figure 3, keeping the other segments fixed. If there are layers which generate smaller costs than the current solution, we record the one giving the largest improvement. Otherwise we store the current assignment. After all segments have been tested, the segments are assigned to their corresponding recorded layers simultaneously. The algorithm therefore propagates depth independent of the order of applied operations. This process is then iterated and terminates if there has not been an improvement of the costs for a fixed number of iterations. The generated solution with lowest costs is returned. Since in hypothesis testing only a small portion of the warped image is changed, we employ an incremental warping procedure as suggested in [5], which also works for our cost function. Furthermore, we only need to test segments if their neighbourhood has changed in the previous iteration.

2.4. Integration

Figure 4 shows the algorithmic integration of the previously described steps. Layer extraction and assignment are iteratively applied. The algorithm terminates if the costs could not be improved for a fixed number of iterations and returns the solution which had the lowest costs.

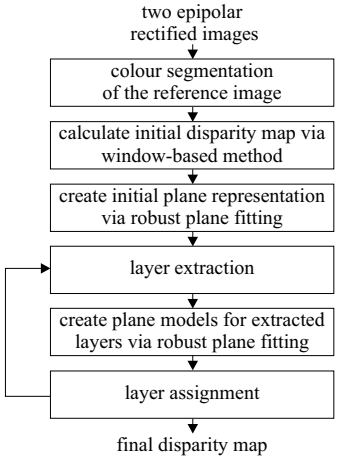


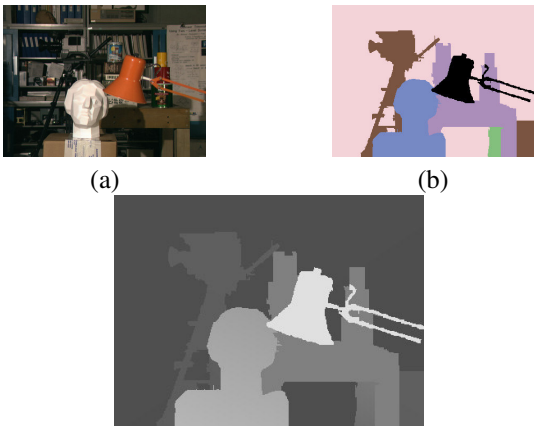
Fig. 4. Algorithmic outline.

3. EXPERIMENTAL RESULTS

We evaluated our algorithm using the test bed proposed by the authors of [1], who compare the performance of 29 different stereo algorithms in the online version of their paper. They provide 4 stereo pairs with corresponding ground truth. For quantitative evaluation the percentage of unoccluded pixels whose absolute disparity error is greater than 1 is used. We applied our algorithm to the test images and submitted the results to the online version of the test bed. At the time of writing this paper our algorithm was ranked as having the second best overall performance among the algorithms tabulated, which demonstrates the high quality of the achieved matching results. The depth maps computed for 3 evaluation pairs are presented in figures 5, 6 and 7. For the computed disparity map in figure 5 the percentage of obtained bad pixels in unoccluded regions is 1.53. For the depth maps in figures 6 and 7 we derive 0.16 and 0.22 percent of wrong pixels. We further tested our algorithm on self-recorded data. Figure 8 shows results for a scene taken in our lab. Since we do not have the ground truth for this image pair, we show a 3D reconstruction of the scene to demonstrate the good quality of the obtained disparity map.

4. CONCLUSION

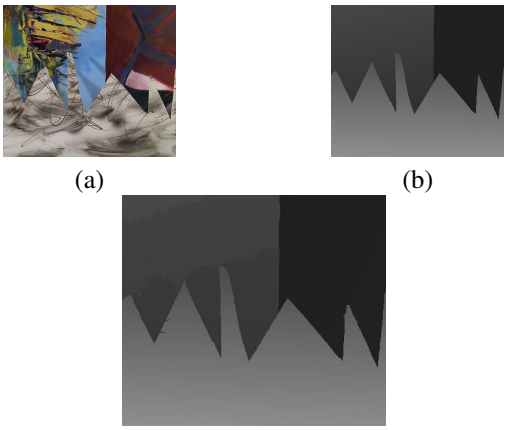
We proposed a new stereo algorithm that uses planar layers to describe the scene. Layers are extracted by a mean-shift-based clustering algorithm. The assignment of segments to layers is made in a hypothesis testing framework. Hypotheses are accepted if they improve the cost function which penalizes occlusions in both views and discontinuities between segments. We demonstrated the performance of the proposed algorithm on images taken from [1] and on self-



(a) (b)

(c)

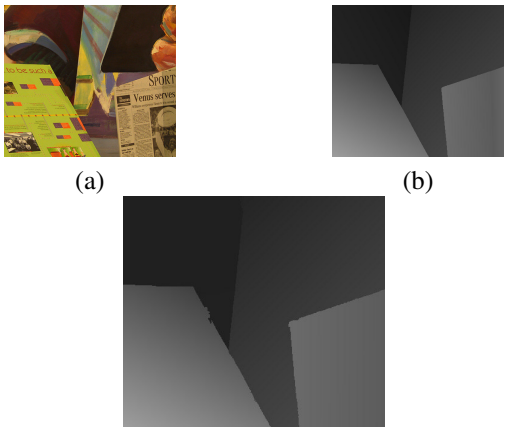
Fig. 5. Results for the Tsukuba dataset. (a) Left image. (b) Final layer assignment. (c) Computed disparity map.



(a) (b)

(c)

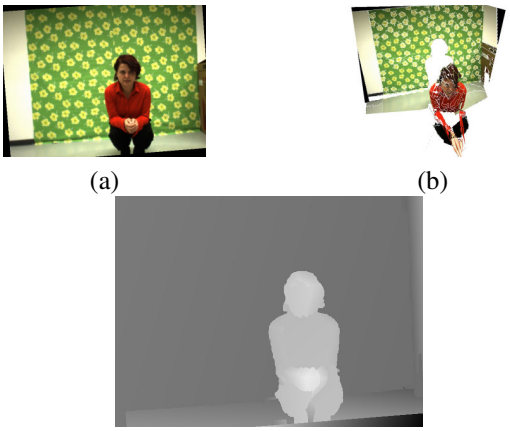
Fig. 7. Results for the Sawtooth dataset. (a) Left image. (b) Ground truth. (c) Computed disparity map.



(a) (b)

(c)

Fig. 6. Results for the Venus dataset. (a) Left image. (b) Ground truth. (c) Computed disparity map.



(a) (b)

(c)

Fig. 8. Results for the self-recorded lab scene. (a) Left image. (b) Reconstructed view. (c) Computed disparity map.

recorded ones. A second rank obtained in the online evaluation on the Middlebury Stereo Vision website confirms the high quality of the achieved results.

5. REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, no. 1/2/3, pp. 7–42, 2002. http://www.middlebury.edu/stereo/.

[2] S. Birchfield and C. Tomasi, "Multiway cut for stereo and motion with slanted surfaces," in *ICCV*, 1999, pp. 489–495.

[3] M. Lin and C. Tomasi, "Surfaces with occlusions from layered stereo," in *CVPR*, 2003, pp. 710–717.

[4] Q. Ke and T. Kanade, "A subspace approach to layer extraction," in *CVPR*, 2001, pp. 255–262.

[5] H. Tao and H. Sawhney, "Global matching criterion and color segmentation based stereo," in *WACV*, 2000, pp. 246–253.

[6] Y. Zhang and C. Kambhamettu, "Stereo matching with segmentation-based cooperation," in *ECCV*, 2002, pp. 556–571.

[7] C. Christoudias, B. Georgescu, and P. Meer, "Synergism in low-level vision," in *ICPR*, 2002, vol. 4, pp. 150–155.

[8] D. Comaniciu and P. Meer, "Distribution free decomposition of multivariate data," *Pattern Analysis and Applications*, vol. 1, no. 2, pp. 22–30, 1999.